Algorithmic Fairness

Year: 2020 | Citations: 414 | Authors: Dana Pessach, E. Shmueli

---

Abstract

An increasing number of decisions regarding the daily lives of human beings are being controlled by artificial intelligence (AI) algorithms in spheres ranging from healthcare, transportation, and education to college admissions, recruitment, provision of loans and many more realms. Since they now touch on many aspects of our lives, it is crucial to develop AI algorithms that are not only accurate but also objective and fair. Recent studies have shown that algorithmic decision-making may be inherently prone to unfairness, even when there is no intention for it. This paper presents an overview of the main concepts of identifying, measuring and improving algorithmic fairness when using AI algorithms. The paper begins by discussing the causes of algorithmic bias and unfairness and the common definitions and measures for fairness. Fairness-enhancing mechanisms are then reviewed and divided into pre-process, in-process and post-process mechanisms. A comprehensive comparison of the mechanisms is then conducted, towards a better understanding of which mechanisms should be used in different scenarios. The paper then describes the most commonly used fairness-related datasets in this field. Finally, the paper ends by reviewing several emerging research sub-fields of algorithmic fairness.